Folding of Subtilisin BPN′: Characterization of a Folding Intermediate†

Jörg Eder, Michael Rheinnecker, and Alan R. Fersht*

MRC Unit for Protein Function and Design, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

Received September 18, 1992; Revised Manuscript Received November 6, 1992

ABSTRACT: Subtilisin BPN′, an extracellular serine protease from *Bacillus amyloliquefaciens*, requires a 77 amino acid pro-sequence for correct folding in vivo. We report the observation of a metastable folding intermediate during the refolding of wild-type and a proteolytically inactive mutant subtilisin BPN′ that lack the pro-sequence. The addition of the pro-sequence as a separate polypeptide chain results in the folding of the intermediate to the native state. The intermediate state of subtilisin is stable at different temperatures, pH values, and salt concentrations for more than a week and retains its competence for folding. The intermediate state possesses a compactness between that of the native and unfolded states. Although it has native-like secondary structure, it shows no distinct near-UV CD spectrum and has a strongly reduced dispersion in the amide and methyl regions of the ^{1}H NMR spectrum. These indicate considerably less tertiary structure than possessed by the native state. However, the intermediate conformation has regions of stable tertiary structure: it has a high-affinity calcium binding site and, after a first noncooperative transition, unfolds with guanidinium chloride in a cooperative process. These results support a folding mechanism for subtilisin BPN′ that comprises a high energy transition state, which is lowered by the interaction with the pro-sequence. The similarity to the folding mechanism of α-lytic protease supports the hypothesis that a common folding mechanism has been developed through convergent evolution.

All the information necessary for the formation of the three-dimensional structure in many proteins is contained in the amino acid sequence (Anfinsen & Scheraga, 1975). As yet, however, the detailed mechanism by which this determination of the tertiary by the primary structure is achieved during the folding process is not understood. Since an immense number of possible conformations are accessible to a polypeptide chain of even moderate length, the random search mechanism of folding is excluded ($t_{1/2} = 10^{50}$ s; Levinthal, 1968; Wetlaufer, 1973), and so nonrandom or "directed" folding pathways must be present. Protein folding studies are thus concerned with what types of intermediate conformations lie between the native and the fully unfolded states on the folding pathway. However, since folding is a cooperative process, small globular proteins are often found to follow a two-state unfolding transition in which only the native and the unfolded polypeptide chains are stable (Kim & Baldwin, 1990).

On the other hand, an unusual folding mechanism of several small globular serine proteases provides a means to circumvent the high cooperativity of the folding process. Subtilisin BPN′, an extracellular serine protease from *Bacillus amyloliquefaciens*, is synthesized as a precurser in form of pre-pro-subtilisin (Wells et al., 1983). The 30 amino acid pre-sequence serves as a signal peptide for protein secretion across the membrane (Wong & Doi, 1986). The pro-sequence, which consists of 77 amino acids, is only transiently connected to the 275 residue protease domain and is required for the correct folding of the enzyme (Zhu et al., 1989). During the folding reaction, the covalent linkage between the pro-sequence and subtilisin is not needed (Zhu et al., 1989). Further, the pro-sequence strongly inhibits the native enzyme (Ohta et al., 1991), suggesting that it functions at a late step on the folding pathway. Folding of subtilisin BPN′ in the absence of the pro-sequence, which does not lead to restoration of enzyme activity (Ikemura et al., 1987), may, however, facilitate the isolation and characterization of a metastable folding intermediate. Two other evolutionarily unrelated serine protease, α-lytic protease (Silen & Agard, 1989) and carboxylpeptidase Y (Winther & Sørensen, 1991), follow a similar folding mechanism. It has been shown recently that during the refolding of α-lytic protease an inactive but folding competent intermediate state can be trapped in the absence of its pro-sequence (Baker et al., 1992).

In this work, we have characterized an intermediate state on, or directly related to, the folding pathway of subtilisin BPN′. The amount of secondary and tertiary structure of the intermediate state has been assessed by several spectroscopic methods and its relative hydrodynamic volume measured compared to that of the native and unfolded polypeptide chain. It has recently been reported (Bryan et al., 1992) that a subtilisin variant, lacking the high-affinity calcium binding site, refolds in the absence of the pro-sequence to the native state. We show here that those studies are consistent with folding to the intermediate state only.

MATERIALS AND METHODS

Materials. Sequanal grade Gdn·HCl,[1] which was used for spectroscopic studies, was purchased from Pierce. Restriction enzymes were from Boehringer Mannheim Corp., Pharmacia, or New England Biolabs. BTFBA was obtained from Lancaster Syntesis Ltd. Factor Xa was from HT Biotechnology Ltd. Methotrexate–agarose was from Sigma. All other reagents were of analytical grade.

Genetic Engineering Methods. Oligonucleotides were synthesized on an Applied Biosystems ABI 380B oligonucleotide synthesizer. The gene sequence encoding subtilisin BPN′ from amino acid 1 to 275 (Thomas et al., 1985) was subcloned via *Sph*I and *Bam*HI restriction sites into the vector

† J.E. was supported by the Deutsche Forschungsgemeinschaft, and M.R. was supported by the MRC and Schering Agrochemicals Ltd.

* To whom correspondence should be addressed.

[1] Abbreviations: BTFBA, 3,5-bis(trifluoromethyl)benzeneboronic acid; CD, circular dichroism; DHFR, dihydrofolate reductase; EDTA, (ethylenedinitrilo)tetraacetic acid; Gdn·HCl, guanidinium chloride; PMSF, phenylmethanesulfonyl fluoride; SDS, sodium dodecyl sulfate.

0006-2960/93/0432-18$04.00/0

pDS56/RBS II (Certa et al., 1986). The *Sph*I and *Bam*HI restriction sites, which were at the 5′ end and 3′ end, respectively, of the subtilisin gene fragment, were generated by PCR (Clackson et al., 1991) with the primers 5′-CGTAGCACATGCGTGCATGCAGTCCGTGCCT-3′ and 5′-CCGACGCCGTTTTGGATCCTTTACTGAGCTGC-3′. In order to replace serine 221 by alanine, site-directed mutagenesis was performed with this plasmid using the primers 5′-CACGTTGCCGGACGCGGCTGCTTTG-3′ and 5′-CGGAGATGCCATGGCCGTACCGTTGTACG-3′ in an inverse PCR reaction (Clackson et al., 1991). The *Eco*RI–*Bam*HI DNA fragment of the mutant clone was then ligated into pTZ18U (Mead et al., 1986).

PCR techniques were also used to subclone the gene fragment corresponding to the pro-sequence of subtilisin BPN′ (Power et al., 1986; Wong & Doi, 1986) into the vector pDS78/RBS II DHFR (kindly provided by D. Stüber). The primers used were 5′-ATGTTAGTCGAAGCTTAGTACGCATGTGCTACGTGATC-3′ and 5′-GTCAGTGAAGATCTGCAGGGAAATCAAACGGGG-3′. The resulting PCR fragment, which had a *Bgl*II restriction site at the 5′ end and a *Hin*dIII restriction site at the 3′ end, was inserted between the *Bgl*II and *Hin*dIII restriction sites of the vector. Hybridization of two complementary oligonucleotides (5′-GATCCCGGGGTTATCGAAGGTCGTATCG-3′ and 5′-GATCCGATACGACCTTCGATAACCCCGG-3′) led to a synthetic DNA fragment that was then ligated into the *Bgl*II site of this plasmid.

Constructs were verified by DNA sequencing (Sanger et al., 1977) of the entire coding sequence. All other recombinant DNA techniques were carried out according to Sambrook et al. (1989).

Protein Expression and Purification. Wild-type subtilisin BPN′ was purified from cultures of *Bacillus subtilis* DB104 harboring the subtilisin gene ligated into plasmid pUB110 as described previously (Thomas et al., 1985). Calcium chloride (2 mM) was added to the culture medium to reduce losses from autolysis.

The S221A subtilisin mutant was expressed in *Escherichia coli* BL21 pLysE cells (Studier & Moffatt, 1986). For the purification of the mutant protein from inclusion bodies, 5.5 g of wet cell pellet was resuspended in 40 mL of 0.05 M potassium phosphate buffer at pH 7.5, containing 1 mM EDTA and 1 mM 2-mercaptoethanol. Cells were ruptured by sonication at 0 °C. After the homogenate was centrifuged at 16500*g* for 15 min, the pellet, which contained predominantly inclusion bodies, was dissolved in a solution of 6 M Gdn·HCl in 0.05 M potassium phosphate buffer, pH 7.5. The solution was dialyzed against 20 mM potassium phosphate buffer, pH 6.4, containing 6 M urea and centrifuged for 30 min at 16000*g*, and the supernatant was loaded onto a column of SP-trisacryl (2.5 cm × 23.5 cm) equilibrated with the same buffer. The S221A mutant protein was eluted with a linear gradient of 20–250 mM potassium phosphate. Fractions that contained >95% pure protein were pooled and dialyzed at a concentration of 0.3 mg of protein/mL against 0.05 M potassium phosphate buffer, pH 6.5. Typically, 15 mg of protein was obtained per liter of cell culture.

Expression of the DHFR-Xa-pro fusion protein was performed at 42 °C in *E. coli* TG2 cells harboring the corresponding expression plasmid. Inclusion bodies were isolated and dissolved as described above. The fusion protein was refolded at a concentration of 0.2 mg of protein/mL by dialysis gainst 0.05 M potassium phosphate buffer, pH 7.0, containing 1 mM 2-mercaptoethanol and 1 mM EDTA. The precipitate was removed by centrifugation. The supernatant was loaded onto a methotrexate–agarose column (1 cm × 20 cm). After the column was washed with the same buffer containing 1 M potassium chloride, the fusion protein was eluted with 0.05 M potassium phosphate buffer, pH 7.0, containing 5 M urea. Fractions, which had DHFR activity (Baccanari et al., 1975), were pooled and dialyzed against 0.05 M Tris-HCl buffer at pH 8.0, containing 0.1 M sodium chloride and 1 mM calcium chloride. Average yields were 21 mg of protein/L of cell culture.

Cleavage of the fusion protein with factor Xa and isolation of the protein was done as described previously (Eder & Kirschner, 1992) except that factor Xa was added to a ratio of $^1/_{1000}$ of the fusion protein by weight. After incubation for 16 h at 4 °C, the protein solution was heated for 3 min at 50 °C to precipitate both DHFR and factor Xa. The supernatant contained the pro-sequence in a pure form. Purified wild-type subtilisin BPN′, S221A mutant subtilisin, and the pro-sequence were stored at −70 °C after dripping the respective protein solutions into liquid nitrogen.

Refolding of Wild-Type and S221A Subtilisin in the Absence of the Pro-Sequence. The proteins were initially unfolded for 2 h at 24 °C in 0.05 M potassium phosphate, pH 4.5, containing 6 M Gdn·HCl or in 25 mM HCl, pH 1.8. Refolding was initiated either by dialysis or by dilution of the unfolded proteins under various conditions.

Unfolded wild-type or S221A mutant subtilisin was dialyzed at 4 °C against 0.05 M potassium phosphate, pH 6.5, or 0.05 M Tris-HCl, pH 7.5 or 8.0. The buffers were supplemented in part with 1 mM $CaCl_2$, 0.5 M $(NH_4)_2SO_4$, or 1 M KCl. Protein concentrations were 3 μM. After dialysis was complete, the protein solutions were incubated at either 4 or 25 °C. PMSF (1 mM) was added to all buffers to avoid autoproteolysis during dialysis of wild-type subtilisin.

Refolding initiated by dilution was carried out on protein solutions at concentrations of 3 μM for samples denatured in HCl and 100 μM for samples denatured in Gdn·HCl. Subtilisin that had been unfolded in Gdn·HCl was diluted 1:100 into the dialysis buffers and incubated at either 4 or 25 °C. Acid-denatured protein was neutralized to pH 7.5 or 8.0 by the addition of Tris base from a 1 M stock solution. Potassium chloride solution (3 M) was then added to a number of the protein samples to a final concentration of 1 M. After refolding by dialysis or dilution, protein solutions were concentrated by ultrafiltration with Amicon PM10 filters prior to performing the measurements described below.

Refolding of Wild-Type and S221A Mutant Subtilisin in the Presence of the Pro-Sequence. For refolding studies in the presence of the pro-sequence, an equimolar mixture of wild-type subtilisin and the pro-sequence was unfolded in a solution of 6 M Gdn·HCl in 0.05 M potassium phosphate, pH 4.5. After incubation for 2 h at 24 °C, the protein solution was dialyzed at 4 °C against 20 mM potassium phosphate buffer at pH 7.0, containing 1 mM $CaCl_2$, 0.5 M $(NH_4)_2SO_4$, and 0.01 mM BTFBA.

The S221A mutant protein was unfolded in a solution of 6 M Gdn·HCl and dialyzed against 0.05 M potassium phosphate, pH 6.5, as described above. The pro-sequence, in 0.05 M potassium phosphate buffer, pH 7.0, containing 0.5 M ammonium sulfate and 1 mM calcium chloride, was then added in an equimolar ratio. The final buffer composition was 0.05 M potassium phosphate, pH 6.8, 0.2 M $(NH_4)_2SO_4$, and 0.5 mM $CaCl_2$. Following these treatments, the protein samples were incubated at 4 °C for the indicated periods of time.

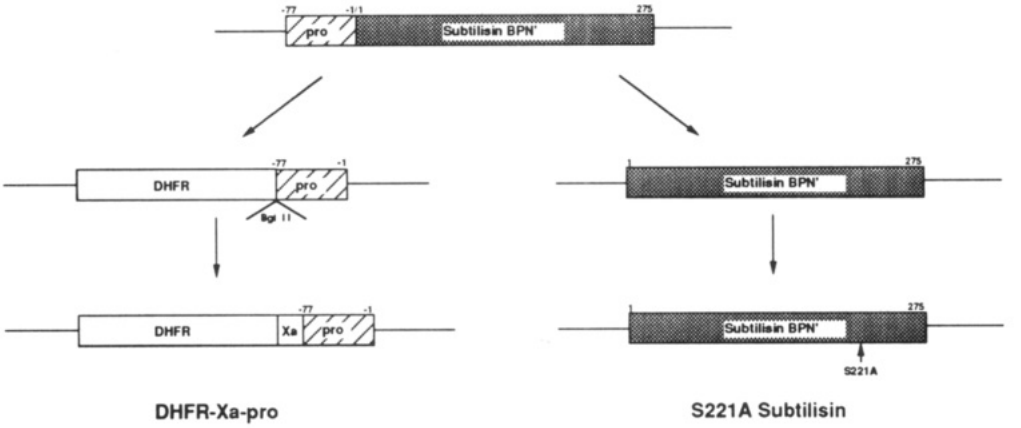

FIGURE 1: Construction of the genes encoding the proteins. The hatched box represents the gene fragment corresponding to the pro-sequence; the shaded box represents the subtilisin BPN′ gene fragment. The number of the first and last amino acid residues of subtilisin BPN′ and the pro-sequence, respectively, are given above the boxes. DHFR-Xa-pro = fusion protein of mouse DHFR and the pro-sequence with linker for factor Xa cleavage. S221A subtilisin = subtilisin BPN′ with serine 221 replaced by alanine.

Determination of Relative Hydrodynamic Volumes. Relative hydrodynamic volumes were measured by gel filtration chromatography on a Superdex 75 column (1 cm × 30 cm) using a Pharmacia FPLC system. The column was equilibrated and eluted with 0.25 M potassium phosphate, pH 6.4, at 22 ± 2 °C. After 100-μL samples of 0.1–0.3 mg of protein/mL were injected, the elution profile was monitored by absorption at 280 nm. The column was calibrated with proteins of known M_r: chymotrypsin inhibitor 2 (7300); bovine milk α-lactalbumin (14 200); mouse DHFR (20 900); bovine erythrocyte carbonic anhydrase (29 000); ovalbumin (45 000); and bovine serum albumin (66 000).

NMR Methods. Protein solutions were concentrated by ultrafiltration with Amicon PM10 filters to concentrations of 0.7–0.8 mM in 50 mM potassium phosphate buffer at pH 6.5, containing 1 mM PMSF. ^{1}H NMR spectra were acquired on a Bruker AMX-500 spectrometer at 15 °C. Spectra were obtained with a 8000-Hz spectral width using 4096 data points. Between 500 and 1000 transients were accumulated per spectrum, depending on the sample concentration.

Circular Dichroism Measurements. CD measurements were carried out with a Jasco J720 instrument at 15 °C. Far-UV CD (200–250 nm) spectra were measured at protein concentrations of 5–7 μM in a 0.1-cm cell. Near-UV CD (250–320 nm) spectra were recorded on solutions of 11–13 μM protein in a 2-cm cell. The spectra, which were averaged over eight accumulations, were corrected for the appropriate (buffer) baseline. Mean residue ellipticity values ($[\theta]_{mrw}$) were calculated using the expression

$$[\theta]_{mrw} = 100\theta_{obs}/lc$$

where θ_{obs} is the observed ellipticity in degrees, c is the molar residue concentration, and l is the light path in centimeters (Schmid, 1989).

For equilibrium unfolding experiments, the ellipticity at 222 nm was measured on protein solutions of 0.08 mg/mL in a 0.1-cm cell. Measurements were made after incubating the protein for 3 h at 20 °C in 0.05 M potassium phosphate buffer, pH 6.4, with the indicated final concentrations of Gdn·HCl. No further change in θ_{222} occurred after longer incubation times.

Fluorescence Measurements. Fluorescence emission spectra were measured with a Perkin Elmer LS 50 luminescence spectrometer at 20 °C with excitation at 280 nm. The slit width of both monochromators was 5 nm. Measurements were carried out with 2.5 μM protein solutions in 0.05 M potassium phosphate buffer at pH 6.4, containing 1 mM PMSF and 1 mM $CaCl_2$.

The binding of calcium ions to the protein was determined by titration with the fluorescent Ca^{2+} chelator Quin2 (Bryant et al., 1985). S221A mutant subtilisin was unfolded in a solution of 6 M Gdn·HCl and refolded by extensive dialysis against 0.05 M potassium phosphate buffer at pH 6.5, containing 22.5 μM Ca^{2+}. For titration experiments, the protein solution (7.5 μM) and the buffer were mixed with a series of aliquots of Quin2 (0.5 mM in Ca^{2+}-free water), respectively. The final sample volumes were 1.0 mL. After the mixtures had been incubated for 20 h in the dark at 4 °C, the Quin2 fluorescence was measured. Measurements were done at 20 °C with excitation at 339 nm (slit width 3 nm) and emission at 500 nm (slit width 8 nm).

RESULTS

Genetic Constructions, Expression, and Purification of the Proteins. The construction of the coding sequences of S221A mutant subtilisin and the DHFR-Xa-pro fusion protein is shown schematically in Figure 1. PCR was used to subclone the gene fragments from plasmid pUB110 (Thomas et al., 1985) into *E. coli* expression vectors. Ligation of the gene fragment corresponding to subtilisin BPN′ from amino acid 1 to 275 into pDS56/RBS II led to a coding sequence of subtilisin with an extra methionine residue at the N-terminus. After introduction of the mutation (S221A) by inverse PCR, the mutant gene was ligated together with the upstream ribosomal binding site of pDS/RBS II into the vector pTZ18U. This vector does not provide a ribosomal binding site downstream of its T7 promoter region. Expression of the S221A mutant gene under the control of the T7 promotor yielded considerable higher expression levels compared to those obtained with the T5 promoter, which is upstream of the ribosomal binding site of ector pDS56/RBS II.

Initial experiments with the two plasmids capable of expressing the pro-sequence by itself were not successful. The crude extracts contained no protein with the expected M_r as detected by SDS–polyacrylamide gel electrophoresis. As an alternative route, we constructed a fusion protein by joining the gene fragment encoding the prosequence to the gene of mouse dihydrofolate reductase (Bujard et al., 1987) as shown in Figure 1. A short synthetic double-stranded DNA fragment, made up of two complementary 28-mer oligonucleotides and coding for a factor Xa cleavage site, was cloned into the *Bgl*II site between the DHFR region and the pro-sequence. The

design of this cleavage site essentially followed the principles outlined by Eder and Kirschner (1992). The decapeptide inserted by this linker fragment, -S-D-P-G-V-I-E-G-R↓-I-G-S-A-G- (the last two amino acid residues of mDHFR and the first two amino acid residues of the pro-sequence are shown in bold type; the factor Xa recognition sequence is underlined), is cleaved by factor Xa as shown by the arrow. This cleavage generates the pro-sequence with three additional amino acids at its N-terminus (IGS-).

The S221A mutant subtilisin and the DHFR-Xa-pro fusion protein were expressed separately in *E. coli*, and the major fraction of each protein was produced in inclusion bodies. However, only expression of the fusion protein at 42 °C led to the formation of inclusion bodies. If cell cultures were induced at 37 °C, most of the overexpressed protein was soluble and considerable proteolytic digestion occurred in the region of the pro-sequence. The inclusion bodies of the fusion protein and the S221A mutant subtilisin were solubilized and purified to >95% purity as judged by SDS–polyacrylamide gel electrophoresis.

Cleavage of the fusion protein by factor Xa was specific. No unspecific cleavage was observed even after prolonged incubation. Heating the protein solution after the Xa digest was complete precipitated both mDHFR and factor Xa. The pro-sequence did not precipitate at this temperature. The molecular weights of the purified proteins were confirmed by mass spectrometry.

Refolding Studies with Wild-Type Subtilisin BPN'. Unfolding of wild-type subtilisin, using concentrated solutions of Gdn·HCl, was carried out at pH 4.5 to minimize the autolytic proteolysis observed when the protein is unfolded at higher pH values (e.g., pH 6.5). At low pH and in 6 M Gdn·HCl, the unfolded subtilisin is stable toward autolytic proteolysis for more than a week as monitored by SDS–polyacrylamide gel electrophoresis. In contrast, wild-type subtilisin, refolded by either dilution or dialysis against 0.05 M potassium phosphate buffer, pH 6.4, undergoes complete autoproteolytic digestion after incubation at 4 °C for 16 h. This proteolysis could be prevented by refolding the protein either in the presence of 1 mM PMSF, which is an irreversible inhibitor for serine proteases, or in the presence of a 10-fold molar excess of chymotrypsin inhibitor 2 (CI2), which inhibits subtilisin BPN' with a K_i of 2.9×10^{-12} M (Longstaff et al., 1990). Since wild-type subtilisin was purified from a *Bacillus* strain deficient in other extracellular proteases, it is likely that at least some of the subtilisin molecules are able to fold to the native conformation. However, the restored activity was found to be too low to be measured under normal subtilisin assay conditions (Longstaff et al., 1990). Additionally, refolding in the presence of CI2 did not yield a subtilisin–CI2 complex as analyzed by gel filtration chromatography (cf. Figure 2). This complex should have been formed in the presence of native subtilisin and eluted from the column with an elution volume of 10.5 mL (data not shown). Its formation has not been observed even after incubation for several days. Thus, although some subtilisin molecules may refold to the native state under these conditions, their amount appears to be too low to be detected either by kinetic activity measurements or by gel filtration chromatography. To avoid the problem of autolytic proteolysis during refolding of subtilisin, folding studies were carried out partly on an almost inactive subtilisin mutant. This variant has Ser 221, which is part of the catalytic triad, replaced by alanine, thereby decreasing the catalytic efficiency by a factor of 2×10^6 (Carter & Wells, 1988).

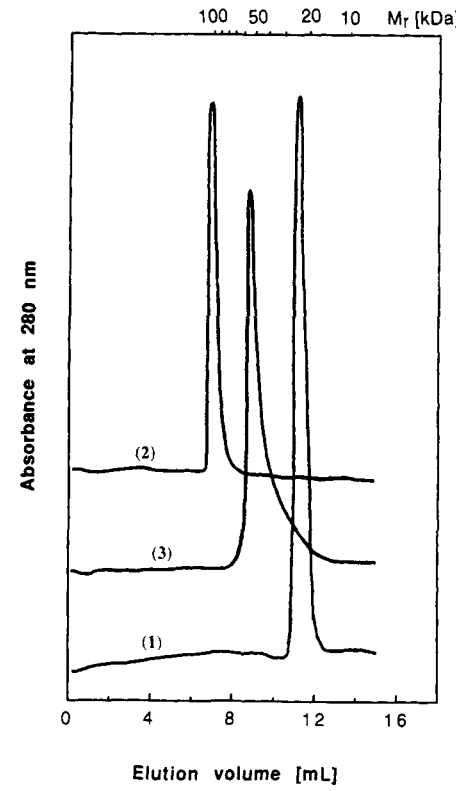

FIGURE 2: Gel filtration chromatography at 22 ● 2 °C. The intermediate state has a hydrodynamic volume between native and unfolded states. (1) Native subtilisin; (2) unfolded subtilisin; (3) subtilisin after refolding in the absence of the pro-sequence. Elution profiles were monitored by absorption at 280 nm at a flow rate of 0.4 mL/min. The elution buffer for the unfolded protein was supplemented with 6 M Gdn·HCl. M_r values on upper ordinate from calibration with marker proteins.

Characterization of a Folding Intermediate. (a) Relative Hydrodynamic Volumes. The relative hydrodynamic volumes of different conformational states of subtilisin were determined by gel filtration chromatography (Figure 2). Native wild-type subtilisin BPN' eluted from the column at a volume of 11.6 mL corresponding to an apparent M_r of 24 000. The calculated M_r of subtilisin is 27 600. By contrast, the apparent M_r of the unfolded protein was 100 000 as determined by gel filtration (elution volume of 6.9 mL). This difference in the elution volume between native and unfolded subtilisin suggests a substantial increase in the hydrodynamic volume of the polypeptide chain upon unfolding.

Both wild-type subtilisin and the S221A mutant protein had the same elution volume of 9.0 mL when refolded by either dilution or dialysis. This value corresponds to an apparent M_r of 52 000. No influence of incubation times between several hours at 24 °C and up to a week at 4 °C was observed. Therefore, refolding of subtilisin occurs only to an intermediate state under these conditions. The compactness of this intermediate state is between that of the native and unfolded states. In addition, the peak corresponding to the intermediate is not symmetric but tails considerably, suggesting that the intermediate state represents an ensemble of different conformations rather than one defined structure.

(b) Circular Dichroism and Fluorescence Spectroscopy. Circular dichroism spectra were measured as an indicator of relative differences in secondary and tertiary structure (Bayley, 1980; Schmid, 1989). Figure 3A shows the far-UV CD spectra of wild-type subtilisin BPN' in its different conformational states. The addition of PMSF to all buffers decreased the signal/noise ratio below 210 nm. The spectrum of the refolded

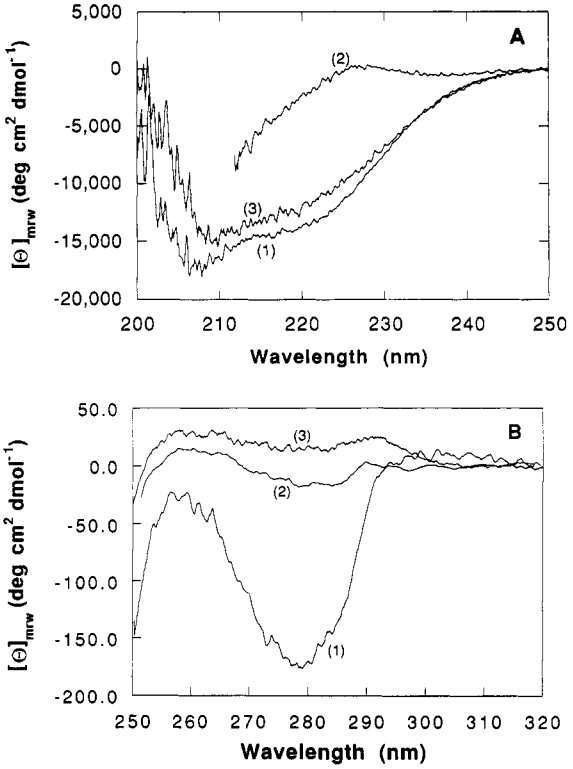

FIGURE 3: CD spectra of different conformations of subtilisin BPN'. (1) Native subtilisin; (2) unfolded subtilisin; (3) subtilisin after refolding in the absence of the pro-sequence. The buffer of the unfolded protein contained 6 M Gdn·HCl. Protein concentrations were 5–13 μM. Spectra were recorded at 15 °C. (A) Far-UV CD spectra measured in a cell with a 0.1-cm path length. (B) Near UV-CD spectra recorded in a cell with 2-cm light path.

protein is very similar to that of native subtilisin. Both spectra, which have ellipticity minima at 208 and 222 nm, reflect well-defined secondary structure. However, the spectrum of refolded subtilisin in its intermediate state has only 90% of the amplitude of the native spectrum at 222 nm. Further, subtilisin BPN' appears to be completely unfolded in 6 M Gdn·HCl, since no residual negative ellipticity signal was observed above 220 nm. The far-UV CD spectra of the unfolded and intermediate states of the S221A mutant protein are identical to that of wild-type subtilisin in its corresponding conformation.

As shown in Figure 3B, the near-UV CD spectra of the native and the intermediate states of subtilisin differ strongly. The spectrum of the native state of wild-type subtilisin BPN' has a minimum in ellipticity at 278 nm, indicating that aromatic amino acid residues are in an asymmetric environment and therefore reflecting a defined tertiary structure. In comparison, the near-UV CD spectra of both the intermediate and the unfolded states show almost no amplitude in this region. The same results were obtained with the S221A mutant protein. Again, extended incubation times (8 days at 4 °C) did not alter the near-UV CD signal.

Fluorescence emission spectra of tryptophan residues in wild-type subtilisin were used to monitor qualitatively any differences in the polarity of their environment upon unfolding and refolding (Schmid, 1989). There are three tryptophans in subtilisin BPN' (W106, W113, W241), and all are located on the surface of the protein, i.e., in part exposed to the solvent (Bott et al., 1988). The fluorescence emission spectrum (Figure 4) of the intermediate state is shifted by 4 nm toward shorter wavelength compared to that of the native state, suggesting a slightly lower degree of exposure of the tryptophan

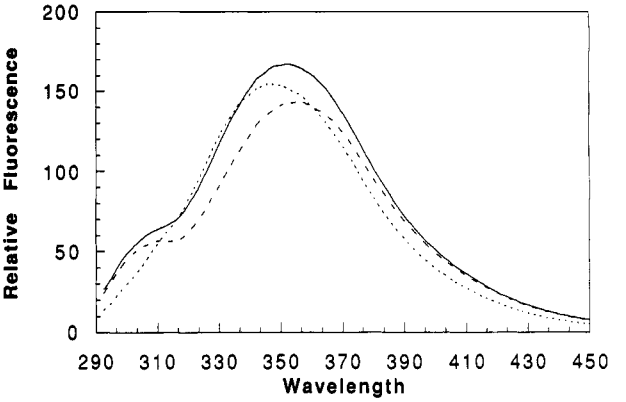

FIGURE 4: Fluorescence emission spectra of different conformations of subtilisin BPN'. Fluorescence was excited at 280 nm. Protein concentrations were 2.5 μM in 0.05 M potassium phosphate, pH 6.5, containing 1 mM PMSF. The temperature was 20 °C. (—) Native subtilisin; (- - -) unfolded subtilisin; (– – –) subtilisin after refolding in the absence of the pro-sequence. Unfolding of subtilisin was by addition of 6 M Gdn·HCl.

residues to the solvent. It is also quenched by about 8%. In contrast, addition of 6 M Gdn·HCl led to a red-shift of the fluorescence spectrum by 4 nm, reflecting complete exposure of the tryptophans to the solvent.

(*c*) *^{1}H NMR Spectra.* ^{1}H NMR spectra of the native and the intermediate states of wild-type subtilisin BPN' were measured to obtain information on differences in tertiary structures independent of aromatic amino acid residues. Figure 5 panels A and B show the spectrum of the native enzyme between 11.5 and 6.0 ppm and 2.0 and –1.0 ppm, respectively. Strong dispersion of the resonances of both the amide protons (between 11 and 7 ppm) and the methyl protons (0–1.5 ppm) indicates well-defined tertiary structure as expected for a protein in the native conformation. Additionally, some α-protons resonate above the water signal between 5.5 and 6 ppm, thus reflecting regions of β-sheet conformation.

In contrast, the intermediate state shows much less dispersion in the amide (Figure 5C) and methyl regions (Figure 5D) of the ^{1}H NMR spectrum. No resonances of α-protons were obtained above 5.5 ppm. Therefore, the intermediate state has either considerable less tertiary structure than the native state or most of its tertiary structure elements are not stable in the time scale of the experiment. The latter would cause resonances to average out during the course of the NMR experiment. No serious problems with aggregation of molecules in the intermediate state were observed under these conditions as indicated by a similar line width of the resonances of the intermediate compared to that of the native state.

(*d*) *Stability of the Intermediate.* Globular proteins with a well-defined tertiary structure frequently unfold in a cooperative transition (Jaenicke, 1991). We used the unfolding transition induced by Gdn·HCl as an additional criterion for assessing the amount of tertiary structure in the intermediate state. Figure 6 shows the equilibrium unfolding transition of the intermediate state of S221A mutant subtilisin measured by its ellipticity at 222 nm. The unfolding followed the gradual and monotonic transition that is typical of noncooperative processes between 0 and 3.4 M Gdn·HCl, accounting for the loss of about 60% of the initial signal. However, ca. 35% of the overall amplitude of the θ_{222} signal decreased cooperatively between 3.4 and 4.0 M Gdn·HCl. These data suggest that although the majority of the secondary structures in the intermediate state unfold independently, there are a few regions of more highly organized secondary structure, i.e., supersecondary structure.

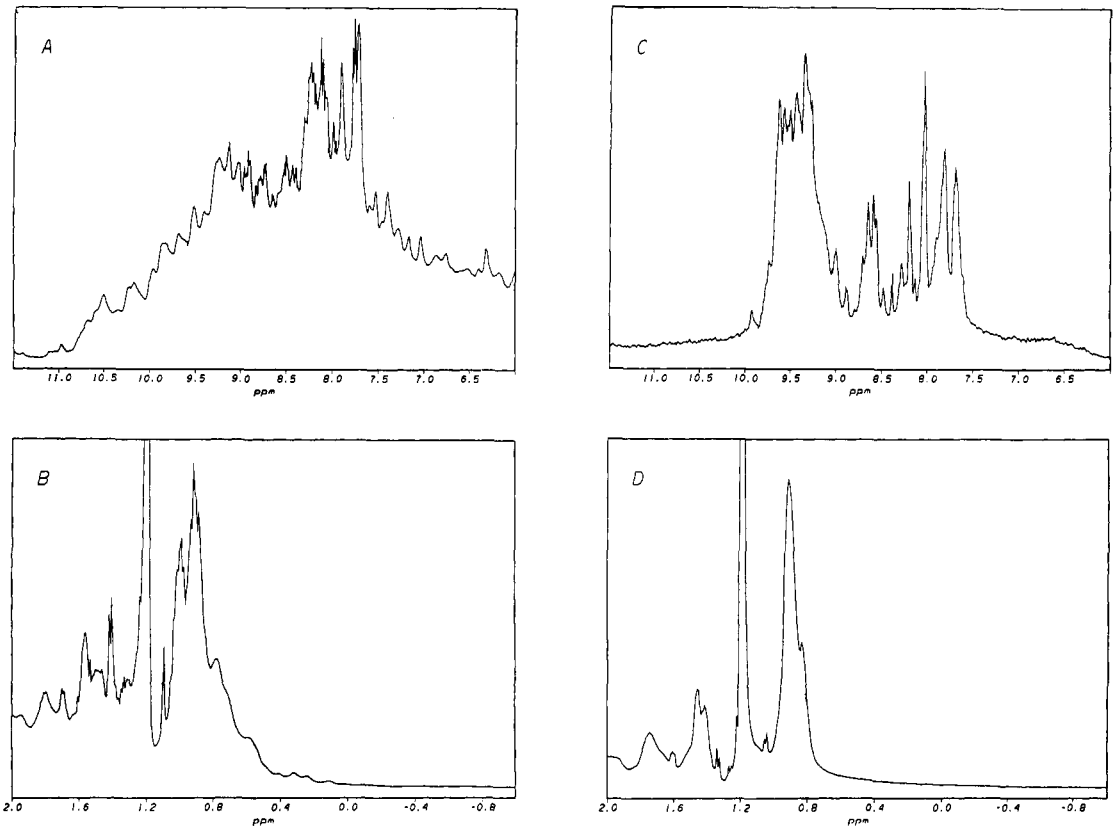

FIGURE 5: 500-MHz ^1H NMR spectra of subtilisin BPN′ in its native and intermediate state. Spectra were acquired at 15 °C in 0.05 M potassium phosphate, pH 6.5, containing 1 mM PMSF after presaturation of the water peak. Protein concentrations were 0.7–0.8 mM. (A) Native state, amide/aromatic region; (B) native state, methyl region; (C) intermediate state, amide/aromatic region; (D) intermediate state, methyl region.

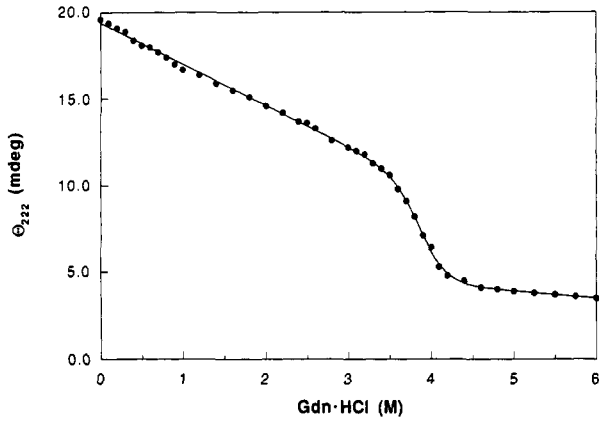

FIGURE 6: Equilibrium unfolding of the intermediate state of S221A mutant subtilisin by guanidinium chloride. Ellipticities at 222 nm were measured in 0.05 M potassium phosphate buffer, pH 6.5, at 20 °C in a cell with a 0.1-cm light path was described under Materials and Methods. Prior to the measurements, the protein was refolded from a concentrated Gdn·HCl solution by dialysis in the absence of the pro-sequence.

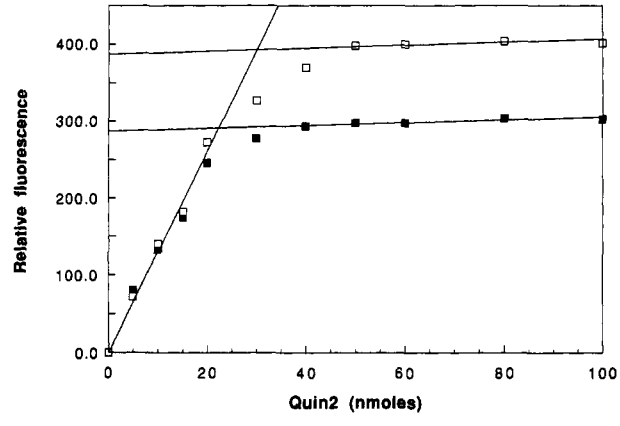

FIGURE 7: Titration of refolded S221A mutant subtilisin (7.5 μM) with aliquots of a solution of the fluorescent Ca^{2+} chelator Quin2 (0.5 mM). The intermediate state has a high-affinity calcium binding site. Prior to the titration, the protein was refolded in the absence of the pro-sequence and dialyzed against 0.05 M potassium phosphate buffer, pH 6.5, containing 22.5 μM calcium. Measurements were at 20 °C with excitation at 339 nm and emission at 500 nm as described under Materials and Methods. The titration points are defined as the points of intersection of the corresponding lines. (Filled squares) Buffer; (open squares) S221A mutant in buffer.

(e) Calcium Binding to the Intermediate. Subtilisin BPN′ has two ion binding sites, one of these (site A) binds Ca^{2+} with high affinity (Bryan et al., 1992). To test the capacity of the intermediate state for calcium binding, we titrated the intermediate state of S221A mutant subtilisin with the fluorescent calcium chelator Quin2. Quin2 binds calcium with a K_a of 1.8×10^8 M^{-1} at pH 7.5 (Linse et al., 1987), whereas the association constant for calcium binding to native subtilisin BPN′ is 7×10^6 M^{-1} Bryan et al., 1992). After the protein solution was dialyzed against a buffer containing 22.5 μM Ca^{2+}, the difference in the titration points between buffer and protein solution was 7.8 ± 0.3 nmol of Quin2 (Figure 7). This value corresponds to a total amount of 7.8 ± 0.3 nmol of calcium bound to the protein or 1.04 ± 0.04 mol of Ca^{2+} bound per mole of protein. The intermediate state, therefore, has a high affinity for calcium ions, underlining the finding that regions of supersecondary structure are already present in this conformation. However, it is still speculative whether the calcium binding site in the intermediate state is identical to that found in the native enzyme.

(f) Influence of Ionic Strength and pH. A variety of different buffers was used to measure the influence of salt or pH on the folding reaction of subtilisin BPN'. No effect of pH on the formation of the intermediate was observed over the range 6.5–8.0. Further, potassium chloride concentrations of up to 1 M, ammonium sulfate concentrations of 0.5 M, and calcium ion concentrations between 0 (in the presence of 0.1 mM Quin2) and 2 mM do not promote further folding of the intermediate state as judged by gel filtration chromatography and near-UV CD spectroscopy. Incubation times were up to several hours at 24 °C or 8 days at 4 °C. However, we did measure an influence of the ionic strength on the folding reaction from the unfolded to the intermediate state. If the acid-denatured S221A mutant subtilisin was refolded by dilution into low-salt buffers (e.g., 0.05 M Tris-HCl, pH 7.5) and incubated for 3 h at 24 °C, two peaks were observed in gel filtration chromatography. These peaks correspond to the unfolded (6.9-mL elution volume) and the intermediate state (9.0-mL elution volume) of the protein (cf. Figure 2). In contrast, if the protein was refolded in buffers containing 1 M potassium chloride and incubated for 1 h at 24 °C, it eluted in one peak (9.0-mL elution volume) corresponding to the intermediate state. Therefore, high ionic strength increases the rate of folding from the unfolded to the intermediate state. Additionally, both increasing ionic strength and increasing temperature were found to cause aggregation of the protein in its intermediate state. Although aggregation was not observed at low salt concentrations, the protein was soluble in buffers containing 1 M potassium chloride only to a concentration of 1.5 and 10 μM at 24 and 4 °C, respectively.

Refolding of Subtilisin BPN' in the Presence of the Pro-Sequence. Unfolded subtilisin can adopt the native state if it is refolded in the presence of the pro-sequence (Zhu et al., 1989). To test the capacity of our pro-sequence as a mediator for folding, wild-type subtilisin BPN' and the pro-sequence were refolded together by dialysis from a solution of 6 M Gdn·HCl. Subtilisin activity was restored during the course of the experiment, indicating the complete folding of the protein molecules to the native state. However, since only a weak subtilisin inhibitor (BTFBA; Philipp & Bender, 1983) was present in the refolded sample, considerable autoproteolysis occurred, thus reducing the yield of recovery of enzyme activity to ca. 1% of the expected value (data not shown).

To examine whether the interaction between subtilisin and the pro-sequence occurs after the intermediate state has formed, a stoichiometric mixture of both S221A subtilisin refolded to the intermediate state and the pro-sequence were incubated at 4 °C. Near-UV CD spectra were measured to detect the formation of tertiary structure induced by the pro-sequence. As shown in Figure 8, only in the presence of the pro-sequence was further folding of the intermediate state observed. It adopts a near-UV CD spectrum qualitatively similar to that of native subtilisin, indicating a massive increase in tertiary structure formation. The reaction appeared to be a slow process. After 8 days of incubation at 4 °C, about 50% of the theoretical amplitude at 278 nm had been restored. However, the yield of signal may have been decreased to some extent by aggregation of the protein due to the long incubation times. The pro-sequence, which has no tryptophan residues, has no detectable near-UV CD signal under these conditions (data to be published elsewhere). The buffer contained calcium ions (0.5 mM) and ammonium sulfate (0.2 M) in order to stabilize the folded structure of the pro-sequence–subtilisin complex. This proved to be necessary since initial experiments with an inactive pro-subtilisin variant showed a strong influence of the salt concentration on the stability of the protein. The ability of native subtilisin to bind CI2 was used as an additional criterion to monitor the folding of the intermediate to the native state in the presence of the pro-sequence. For this analysis, the above reaction mixture was supplemented with a 10-fold molar excess of CI2 and analyzed by gel filtration chromatography. Beside the peak corresponding to unbound CI2 (14.2-mL elution volume), two other peaks eluted from the column in a ratio of 2:8. The elution volumes of these two peaks were 9.0 and 10.5 mL (cf. Figure 2), respectively, thus corresponding to the apparent M_r values of the intermediate state (52 000) and of the subtilisin–CI2 complex (34 900). SDS–polyacrylamide gel electrophoresis revealed that while both peaks contained subtilisin, the second one contained also CI2 (data not shown). Therefore, complex formation between both proteins occurred.

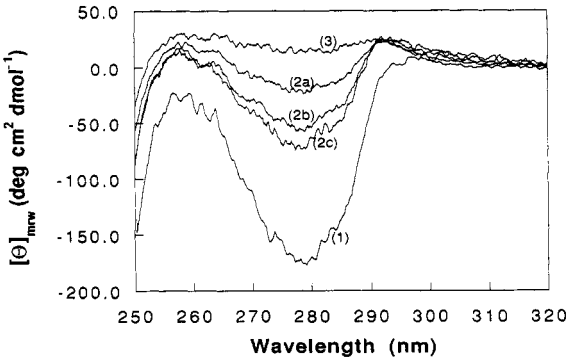

FIGURE 8: Refolding of S221A mutant subtilisin in the presence of the pro-sequence. Near-UV CD spectra recorded in a cell of 2-cm light path at 15 °C. Subtilisin concentrations were 5 μM in 0.05 M potassium phosphate, pH 6.8, 0.2 M $(NH_4)_2SO_4$, and 0.5 mM $CaCl_2$. (1) Native wild-type subtilisin BPN'. (2a), (2b), and (2c) are S221A mutant subtilisin refolded by dialysis in the presence of 5 μM pro-sequence after 1, 4, and 8 days of incubation time, respectively. (3) Wild-type subtilisin BPN' refolded in the absence of pro-sequence after 8 days of incubation time (spectra 1 and 3 are taken from Figure 3B).

DISCUSSION

Subtilisin BPN' is synthesized as a precurser containing a 77 amino acid pro-sequence, which is transiently required for the correct folding of the protease domain (Zhu et al., 1989). The absence of the pro-sequence in the refolding reaction of subtilisin has made it possible to trap an inactive, but folding-competent, state. Structural characterization shows this intermediate state to have properties different from that of the native and unfolded states.

Although the far-UV CD spectrum of the intermediate state implies a content of secondary structure similar to that of the native state, little or no organized tertiary structure was found as judged by the absence of a significant near-UV CD amplitude and a substantial loss in dispersion in the amide and methyl regions of its ^1H NMR spectrum. Further, the intermediate state, which is considerably more compact than the unfolded polypeptide chain, has a less compact conformation than one would expect for a small globular protein with defined tertiary structure. All these structural properties of the intermediate state point toward an intermediate conformation frequently found for other proteins and termed "molten globule" (Dolgikh et al., 1984; Kuwajima, 1989) or "collapsed intermediate" (Kim & Baldwin, 1990). This common intermediate may be characterized as a nonspecific assembly of secondary structure segments brought about by nonspecific hydrophobic interactions. However, a partially

cooperative unfolding transition of subtilisin in its intermediate state suggests the presence of localized regions of supersecondary structure. Such regions, also found in the intermediate state of α-lactalbumin (Kuwajima, 1989), are believed to act as folding initiation sites for the formation of tertiary structure. The unusual stability of the intermediate state of subtilisin BPN′ under conditions in which the native protein is also stable may facilitate a more detailed structural characterization, e.g., by multidimensional NMR analysis, to identify these important structural motifs.

The stability of the intermediate state, however, its reduced substantially in the presence of the pro-sequence. Thus, the protein molecules fold to the energetically more favorable native conformation as indicated by the increase in their near-UV CD spectrum and their final capacity to bind chymotrypsin inhibitor 2. Since the folding competence of the intermediate state is retained also in the absence of the pro-sequence, it is either directly located on the folding pathway or in equilibrium with a conformation on the folding pathway. The apparent high stabilities of the native as well as the intermediate states in the absence of the pro-sequence appear to result from a high-energy transition state between the two conformations. Stabilization of this transition state mediated by the interaction with the pro-sequence is therefore likely to promote folding of the intermediate to the native conformation.

Although pro-subtilisin refolds fast to produce active subtilisin (Ikemura & Inouye, 1988), the folding reaction of the intermediate to the native state in the presence of the pro-sequence as separate polypeptide chain is slow (cf. Figure 8). The simplest explanation for this discrepancy is the order of the respective reactions, since a reaction is generally expected to be faster as an intramolecular than as an intermolecular process. Therefore, to estimate the height of the energy barrier between the native and the intermediate states by comparing the rates of folding in the absence and presence of the pro-sequence, one has to compare the folding rates of subtilisin and pro-subtilisin. Measurements with an inactive pro-subtilisin variant are currently in progress.

It has been reported recently that a variant of subtilisin BPN′, in which the high-affinity calcium binding site has been deleted, can refold to its native state in the absence of the pro-sequence (Bryan et al., 1992). The rate of folding was found to be substantially accelerated under conditions of high ionic strength. These findings are in marked contrast to the above data and require interpretation. Although we observed a strong influence of the ionic strength on the folding reaction from the unfolded to the intermediate state, we were unable to detect any conversion of the intermediate to the native state at high ionic strength in the absence of the pro-sequence. Even under conditions at which the subtilisin variant was found to be fully refolded in less than 4 min (Bryan et al., 1992), no measurable amount of native enzyme was detected in our study after several days. The conclusions drawn earlier (Bryan et al., 1992) are based mainly on far-UV CD and fluorescence spectroscopic studies, which we have now found cannot distinguish unambiguously between the native and the intermediate states of subtilisin (see Figures 3A and 4). It is thus possible that only the folding reaction from the unfolded to the intermediate state was observed by Bryan et al. (1992). Alternatively, the high-affinity calcium binding site of subtilisin, likely to exist already in the intermediate structure, may prevent it from folding to the native state. However, this hypothesis is somewhat unlikely, since the presence of calcium ions normally accelerates the folding reaction of proteins containing a calcium binding site (Kuwajima et al., 1988, 1989). Further, we did not measure any conversion of the intermediate to the native conformation in the absence of calcium, which could be expected if calcium binding prevents further folding. It would therefore be interesting to learn from additional experiments on the folding of the calcium free variant of subtilisin BPN′ whether it is indeed capable of refolding to the native state.

The structural properties of the intermediate state of subtilisin BPN′ are remarkably similar to those of the intermediate state reported previously for α-lytic protease (Baker et al., 1992). α-Lytic protease belongs to a different class of serine proteases (Creighton, 1985) and is unrelated in structure and sequence. α-Lytic protease and subtilisin have developed an identical catalytic mechanism through convergent evolution. The requirement for the pro-sequence for correct folding of the respective protease domain and the existence of a common intermediate state, which is formed on the folding pathway, demonstrates that also a similar folding mechanism has been developed during evolution. Further, the widespread use of pro-sequences by extracellular bacterial and other serine proteases (Wells et al., 1983; Silen & Agard, 1989; Barr, 1990; Winther & Sørensen, 1991;) suggests that this folding mechanism may be a common feature of this type of enzyme. Extracellular, nonspecific proteases need to be extremely stable toward both unfolding and (auto)proteolysis. One way to gain such a stability is to have a high-energy barrier on the folding pathway between the native and other less stable conformations. The role of the pro-sequence thus appears to be to give the protease access to the final stable and active conformation.

ACKNOWLEDGMENT

We thank M. Bycroft for performing the NMR experiments, P. Skelton for carrying out the mass spectroscopy, and T. Gray and T. Buss for synthesizing the oligonucleotides. We appreciate critical comments from A. G. Day.

REFERENCES

Anfinsen, C. B., & Scheraga, H. A. (1975) *Adv. Protein Chem. 29*, 205–300.

Baccanari, D. P., Phillips, A., Smith, S., Sinski, D., & Burchall, J. (1975) *Biochemistry 14*, 5267–5273.

Baker, D., Sohl, J. L., & Agard, D. A. (1992) *Nature 356*, 263–265.

Barr, P. J. (1990) *Cell 66*, 1–3.

Bayley, P. M. (1980) in *An Introduction to Spectroscopy for Biochemists* (Brown, S. B., Ed.) pp 148–234, Academic Press, London.

Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B., & Powers, S. (1988) *J. Biol. Chem. 263*, 7895–7906.

Bryan, P., Alexander, P., Strausberg, S., Schwarz, F., Lan, W., Gilliland, G., & Gallagher, D. T. (1992) *Biochemistry 31*, 4937–4945.

Bryant, D. T. W. (1985) *Biochem. J. 226*, 613–616.

Bujard, H., Gentz, R., Lanzer, M., Stüber, D., Müller, M., Ibrahimi, I., Häuptle, M.-T., & Dobberstein, B. (1987) *Methods Enzymol. 155*, 416–433.

Carter, P., & Wells, J. A. (1988) *Nature 332*, 564–568.

Certa, U., Bannwarth, W., Stüber, D., Gentz, R., Lanzer, M., LeGrice, S., Guillot, F., Wendler, I., Hunsmann, G., Bujard, H., & Mous, J. (1986) *EMBO J. 5*, 3051–3056.

Clackson, T., Güssow, D., & Jones, P. T. (1991) in *PCR: A Practical Approach* (McPherson, M. J., Quirke, P., & Taylor, G. R., Eds.) pp 187–214, IRL Press, Oxford.

Creighton, T. E. (1985) *J. Phys. Chem. 89*, 2452–2459.

Dolgikh, D. A., Kolomiets, A. P., Bolotina, I. A., & Ptitsyn, O. B. (1984) *FEBS Lett. 165*, 88–92.

Eder, J., & Kirschner, K. (1992) *Biochemistry 31*, 3617–3625.
Ikemura, H., & Inouye, M. (1988) *J. Biol. Chem. 263*, 12959–12963.
Ikemura, H., Takagi, H., & Inouye, M. (1987) *J. Biol. Chem. 272*, 7859–7864.
Jaenicke, R. (1991) *Biochemistry 30*, 3147–3161.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631–660.
Kuwajima, K. (1989) *Proteins 6*, 87–103.
Kuwajima, K., Sakuraoka, A., Fueki, S., Yoneyama, M., & Sugai, S. (1988) *Biochemistry 27*, 7419–7428.
Kuwajima, K., Mitani, M., & Sugai, S. (1989) *J. Mol. Biol. 206*, 547–561.
Levinthal, C. (1968) *J. Chim. Phys. Phys.-Chim. Biol. 65*, 44–45.
Linse, S., Brodlin, P., Drakenberg, T., Thulin, E., Sellers, P., Elmden, K., Grunström, T., & Forsen, S. (1987) *Biochemistry 26*, 6723–6753.
Longstaff, C., Campbell, A. F., & Fersht, A. R. (1990) *Biochemistry 29*, 7339–7347.
Lorimer, G. H. (1992) *Curr. Opin. Struct. Biol. 2*, 26–34.
Mead, D. A., Szczesna-Skorupa, E., & Kemper, B. (1986) *Protein Eng. 1*, 67–74.
Ohta, Y., Hojo, H., Aimoto, S., Tobayashi, T., Zhu, X., Jordan, F., & Inouye, M. (1991) *Mol. Microbiol. 5*, 1507–1510.
Philipp, M., & Bender, M. L. (1983) *Mol. Cell. Biochem. 51*, 5–32.
Power, S. D., Adams, R. M., & Wells, J. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3096–3100.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Habor, NY.
Sanger, F., Nicklen, S., & Coulson, A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schmid, F. X. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 251–285, IRL Press, Oxford.
Silen, J. L., & Agard, D. A. (1989) *Nature 341*, 462–464.
Studier, F. W., & Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113–130.
Thomas, P. G., Russel, A. J., & Fersht, A. R. (1987) *Nature 318*, 375–376.
Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., & Chen, E. Y. (1983) *Nucleic Acids Res. 11*, 7911–7925.
Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 697–701.
Winther, J. R., & Sørensen, P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9330–9334.
Wong, S. L., & Doi, R. H. (1986) *J. Biol. Chem. 261*, 10176–10181.
Zhu, X., Ohta, Y., Jordan, F., & Inouye, M. (1989) *Nature 339*, 483–484.